A test for anchoring and yea-saying in experimental consumption data

van Soest, A.H.O.; Hurd, M.

Published in:
Journal of the American Statistical Association

Publication date:
2008

A Test for Anchoring and Yea-Saying in Experimental Consumption Data*

Arthur van Soest, RAND and Tilburg University
and Michael Hurd, RAND

This version: November 2006

Abstract

We analyze experimental survey data, with a random split into respondents who get an open-ended question on the amount of total family consumption - with follow-up unfolding brackets (of the form: is consumption $X or more?) for those who answer "don't know" or "refuse" - and respondents who are immediately directed to unfolding brackets. In both cases, the entry point of the unfolding bracket sequence is randomized. Allowing for any type of selection into answering the open-ended or bracket questions, a nonparametric test is developed for errors in the answers to the first bracket question that are different from the usual reporting errors that will also affect open-ended answers. Two types of errors are considered explicitly: anchoring and yea-saying. We reject the joint hypothesis of no anchoring and no yea-saying. Once yea-saying is taken into account, we find no evidence of anchoring at the entry point.

JEL classification: C81, D12 Key words: Unfolding brackets, Non-response, Anchoring, Yea-saying

Author contacts:

Arthur van Soest, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands.

Phone: +31-13-4662028; fax: +31-13-4663280; email: avas@uvt.nl

Michael Hurd, RAND, P.O. Box 2138, Santa Monica, CA 90407-2138, U.S.A.

Phone: +1-310-3930411 ext. 6945; fax: +1-310-4516923; email: mhurd@rand.org

*Financial support was provided by the National Institute of Aging. The comments of three anonymous referees, an associate editor, the editor, Arie Kapteyn and seminar participants at Bonn University, USC, UC Davis, and Tilburg University are gratefully acknowledged.

1 Introduction

Item non-response in survey data is a well-known problem, particularly if questions relate to sensitive information that respondents are not willing to provide or to information that respondents do not know exactly and find hard to estimate. Examples are the values of household assets like saving accounts or stocks and bonds, or the value of total family income or consumption. Item non-response becomes particularly problematic if the information is not missing at random, i.e., if the probability of not responding correlates with the amount in question conditional on a set of covariates. See, e.g., Manski (1989, 1995) and Little and Rubin (2002, Chapter 15).

Follow-up questions are often used to reduce the loss of information due to item non-response by extracting partial, categorical, information from initial non-respondents (cf., e.g., Juster and Smith, 1997). Particularly in telephone interviews where it is difficult to show range cards on which respondents can choose in one step from a larger number of categories, unfolding brackets can be used to collect banded information. For example, a respondent who answers “don’t know” or “refuse” to a question on total family consumption in the past month then gets the question “Can you say whether it was \$2000 or more?” If the answer is affirmative, the next question is “Was it \$5000 or more?” etc.

The majority of initial non-respondents typically appear to answer one or more unfolding bracket questions, implying that unfolding brackets indeed lead to additional information. Several experiments, however, have shown that the distribution of the categorical answers obtained in an unfolding bracket design depends on the order in which the bracket points are presented. For example, if the first question is “Was it \$ 2000 or more?” i.e., the entry point is \$2000, the distribution gives more weight to the high consumption outcomes than if the entry point were \$500. A psychological explanation is that if people are unsure about the exact amount, the entry point serves as an anchor that provides some information about this amount. This phenomenon is known as anchoring (Tversky and Kahneman, 1974).

The literature shows that anchoring exists and becomes more prominent the more uncertain the respondent is about the exact answer (see, e.g., Jacowitz and Kahneman, 1995). In order to use the answers to unfolding bracket questions for meaningful analysis, the possibility of an anchoring bias needs to be taken into account. Several models for what the answers to bracket questions look like if they suffer from anchoring have been introduced and estimated, using different sources of experimental data with random entry points. Examples are Hurd et al. (1998), Herriges and Shogren (1996), and Cameron and Quiggin (1994). A comparison between competing models for anchoring, however, has not yet been performed.

A second source of incorrect answers in an unfolding bracket design is acquiescence or "yea- saying," the phenomenon that people have a tendency to answer yes rather than no (Schuman and Presser, 1996; Hurd, 1999). For example, the number of "yes" answers to the question "Was it $2000 or more?" will be higher than the number of "no" answers to the question "Was it less than $2000?" The design of the data used in our experiment makes the answers vulnerable to yea-saying.

Both anchoring and yea-saying are framing effects in the sense of Schwarz et al. (1985). Bracket questions may induce uncertain respondents to adjust their expectations and give an answer that does not correspond to the estimate they had before going into brackets. Strack and Mussweiler (1997) conjecture that there would be no anchoring if subjects had absolutely certain knowledge. Schwarz et al. (1985, p. 394) advise to use open-ended answers to avoid this kind of framing effects, but this has the drawback of substantial non-response as we discussed above. None of this implies that open-ended answers are error-free - they reflect the respondents' initial point estimates of the amount in question and can themselves suffer from underreporting, focal point bias, etc. The issue here is whether bracket questions induce new errors, implying that a bracket answer does not always correspond to the respondent's estimate that would be reported in an open-ended answer.

In this paper, we do not analyze reporting errors in open-ended questions but focus on testing for anchoring and acquiescence errors induced by the bracket questions. We use experimental data on household consumption from the AHEAD 1995 survey described in Hurd et al. (1998) in which the sample is randomly split in respondents starting with an open-ended question and respondents who immediately go into an unfolding bracket sequence with random entry point. We first develop a test for the null hypothesis that answers to the first bracket question correspond exactly to the respondents' estimates given in open-ended answers, allowing for any form of selective response behavior, i.e., for an arbitrary correlation between whether an open-ended answer or bracket answer is given and the respondent's best estimate of the consumption amount. We find that this null hypothesis is rejected. We then show that a very simple model of yea-saying can explain why the hypothesis is rejected, suggesting that yea-saying rather than anchoring is the source of the problem. We also show that yea-saying correlates with characteristics in a way that is in line with the notion that more uncertainty leads to more yea-saying.

This paper is organized as follows. In the next section, the existing literature of anchoring and yea-saying is summarized. In Section 3, we describe the experimental data. Section 4 explains the non-parametric test when full non-respondents - those who neither give an open-ended answer, nor a bracket answer - are discarded. The test is applied in Section 5. Section 6 takes the full non-respondents into account, using the notion of worst-case bounds, as in Manski (1989). Section 7 incorporates simple ways of "fixed probability" anchoring and acquiescence, to try and disentangle the effects of anchoring and yea-saying. This section also presents some estimates of yea-saying for different groups, defined by, for example, education level or memory skills. Section 8 concludes.

2 Anchoring Models and Acquiescence Bias

Psychological experiments provide extensive evidence of subjects' cognitive illusion when dealing with uncertainty (Rabin, 1998), such as context effects - effects produced by the nature of the questionnaire such as the wordings of questions or the order in which they are presented (Sudman, Bradburn and Schwarz, 1996). A prominent example is anchoring - the phenomenon that responses are influenced by cues contained in the questions. A psychological explanation is that the respondent treats the question as a problem solving task and uses the information in the question as a cue, seeking a framework to construct a correct solution. Using the information in the question may induce respondents to update their initial (point) estimate. This should be more important if the respondent is more uncertain or knows less about the topic, but evidence of anchoring has even been found among experts (Northcraft and Neale, 1987).

An intuitively appealing anchoring model is the fixed-point model used by Herriges and Shogren (1996), O'Connor et al. (1999), and Hurd et al. (2001). It assumes that respondents update their beliefs about an unknown amount if new information arrives. In this model, the entry point B_1 (the amount mentioned in the first bracket question) serves as an anchor for follow-up questions. The respondents have an initial point estimate $\hat{Y}$ that can be interpreted as a mean or other location measure of the respondent's prior distribution. In the second bracket question, respondents do not compare the bracket point B_2 given in the question to $\hat{Y}$, but to their updated estimate, a weighted mean $\hat{Y}^* = (1-\gamma)\hat{Y} + \gamma B_1$, for some γ

between 0 and 1. The intuition is that a respondent who is uncertain about the true amount Y, sees the entry point B_1 as informative about Y and adjusts the initial estimate towards B_1. Herriges and Shogren (1996) apply their model to data on willingness to pay for water quality improvement and find an estimate for γ of 0.36, with standard error 0.14. They also discuss the possibility that γ varies with $\hat{Y}$ (or Y) and B_1. An important feature of the fixed-point model is that answers to the first bracket question are not affected: although the respondent's initial estimate of the true amount will be drawn towards the entry point, it remains on the same side.

Results of Jacowitz and Kahneman (1995) and Green et al. (1998) suggest that respondents already give biased answers to the first bracket question. For estimates of objective quantities as well as willingness to pay for public goods, they find that, for high entry point values, the proportion of bracket respondents who report that the amount is larger than the entry point exceeds the fraction of open-ended answers larger than the entry point.

Hurd et al. (1998) specify a parametric model that captures this phenomenon in a symmetric way, biasing probabilities smaller than 0.5 upwards and probabilities larger than 0.5 downward. In this model respondents do not compare their initial estimate $\hat{Y}$ to the entry point B_1, but instead compare $\hat{Y}+\epsilon_1$ to B_1, where ϵ is a mean zero error term, assumed to be normally distributed and independent of other components of the model. Hurd et al. (1998) use the same device also at follow-up bracket questions, with independent errors that can have different variances. They call their model a gating model, in which respondents have to pass a number of gates to reach their final bracket answer. At each gate k, new noise is introduced, in the form of an error term ϵ_k, independent of $\hat{Y}$ and previous errors $\epsilon_1, \ldots, \epsilon_{k-1}$.

Somewhat similar, though not yet applied in this context, are the models for binary regression with contaminated data (e.g., Copas, 1988) or misclassification (e.g. Hausman et al., 1998). Assume that with some fixed "gating" probability $P[Gat]$, people give an answer that does not correspond to their initial estimate $\hat{Y}$. For answers to the question "Is the amount B_1 or more?" this implies $P[\text{“}no\text{”}] = P[\hat{Y} < B_1](1 - P[Gat]) + P[\hat{Y} \geq B_1]P[Gat]$ and $P[\text{“}yes\text{”}] = P[\hat{Y} \geq B_1](1 - P[Gat]) + P[\hat{Y} < B_1]P[Gat]$. If $P[Gat]$ does not depend on $\hat{Y}$ and is thus the same whether the initial estimate of the true amount is larger or smaller than the entry point, the reported probability that the amount is at least B_1 will be smaller than $P[\hat{Y} \geq B_1]$ if $P[\hat{Y} \geq B_1] > 0.5$ and larger than $P[\hat{Y} \geq B_1]$ if $P[\hat{Y} \geq B_1] < 0.5$.

Yea-saying or acquiescence is a well-known phenomenon with a long history in cognitive psychology. It refers to a tendency to agree with statements but also extends to yes/no questions (Jackman, 1973). Schuman and Presser (1996, Chapter 8) give an overview of theoretical explanations in the psychology literature. The first is a personality trait, that can be treated like any other personality trait in psychology (Couch and Keniston, 1960). The second characterizes acquiescence as a form of deference shown by low-status respondents towards interviewing staff (Lenski and Leggett, 1960). The third sees acquiescence as the tendency of poorly educated people with low cognitive abilities to be uncritical of statements made by the interviewers (Campbell et al., 1960). Similar to anchoring, it can be expected that acquiescence is more prevalent if the questions are less precise or the respondent is more uncertain about the correct answer.

A direct test of yea-saying on an economic issue is performed by Hurd (1999). He uses experimental data from the 1996 wave of the Health and Retirement Study, a representative sample of the US population aged 54-64 with their spouses. Respondents were asked "About how much could you sell your home for in today's housing market?" They were forced into brackets and randomly assigned to an entry point E (\$50,000, \$100,000 or \$150,000) and to one of three questions: "Would it be more than E," "Would it be E or more?" or "Would it be less than E or E or more?" Hurd found that the first two (unbalanced) formats led to many more "yes" answers than the third (balanced) format for non-financial respondents, while there was hardly any difference for financial respondents (i.e., the person in the household most knowledgeable in financial matters). He interpreted this as evidence of acquiescence related to uncertainty.

Acquiescence bias is also discussed in the willingness-to-pay (WTP) literature. Boyle et al. (1998) find evidence of yea-saying for bid levels in the upper tail by comparing open-ended answers and bracket answers on WTP for a moose hunting site. On the other hand, Frykblom and Shogren (2000) used experimental data to compare open-ended and discrete choice answers on students' WTP for a specific book and found no differences.

3 The AHEAD Wave 2 Consumption Experiment

The data we used are described in detail in Hurd et al. (1998). The AHEAD panel (Assets and Health Dynamics Among the Oldest Old) is roughly representative for the US population over 70 and their spouses. In the descriptive statistics and estimations throughout this paper, we use the sample weights and primary sampling units provided with the public use file to account for the multi-level design features of the data (see the documentation on the HRS web site). The second wave was done in 1995, mainly by telephone interviews. At the end of the regular survey, regular AHEAD participants were asked to complete an additional experimental module. In case of proxy interviews (901 interviews with a relative or other person answering the questions on behalf of a respondent who is unable to do so), the experimental module was skipped. Of the respondents asked, about 80% were willing to do the experimental module and attained the questions on consumption (a few dropped out during the module before getting to the consumption questions). This is the sample used by Hurd et al. (1998) and for our study. We do not address selection into participation in the experimental module. Thus our findings refer to respondents willing and able to sit through a long interview including the experimental module and are not necessarily representative for the complete US population of 70 and older.

The participants in the experimental module were randomly assigned to one of seven groups of approximately equal size. All groups got some questions on household consumption, but the question format differed across the groups. Three groups (sample I) started with an open question: "How much did your household spend on consumption in the past month?" Those who answered "don't know" or "refuse" then got unfolding bracket questions, with a different entry point E for each group. The first unfolding bracket question was: "Would the amount be \$$E$ or more in the past month?" with $E = 500$, $E = 1000$ or $E = 2000$. If a respondent answered "don't know" or "refuse" to a bracket question, the sequence was stopped. The other four groups (sample II) immediately were given bracket questions, with different entry points for the four groups (\$500, \$1000, \$2000 and \$5000).

This sample consists of 4885 observations. About 2.5% of them did not give an answer to either an open-ended question or a follow-up or direct bracket question, see Appendix A; these observations are discarded for now (but will be incorporated in Section 6). This leaves 4759 observations. About 44.5% of these (sample I) started with an open question. Almost two thirds of them gave an open-ended answer (1416 observations). The remaining 681 gave at least one bracket answer. The first panel of Table 1 presents the fractions with consumption less than each of the entry point values in the open-ended answers and according to the follow-up bracket answers, separately for the three groups in sample I with different entry points. There are several explanations for differences between open-ended answers and bracket answers and between bracket answers with different entry points. High consumption families could be more concerned about their privacy and less willing to give a precise number, i.e., open-ended answers could under-represent high amounts. Moreover, respondents may make different mistakes in open-ended answers and bracket answers. The differences between the distributions of bracket responses for the three entry points suggest that some bracket answers are affected by the (randomly assigned) entry point, but significance levels are rather low due to the small numbers of observations.

The bottom panel of Table 1 shows the distributions for the immediate bracket respondents by entry point. A salient feature is that the distribution shifts with the entry point. In this case the numbers of observations are much larger and most of the differences are significant. The studies discussed in the previous section provide several explanations for these differences. Fixed-point anchoring affects the answers to bracket questions from the second bracket question onwards, with the entry point in the first bracket question acting as an anchor. This makes intuitive sense and is well-established in the literature, which is why it is not the issue of the current study. Fixed-point anchoring does not affect the answer to the first bracket question. In this paper we focus on errors in answers to the first bracket question. We do not analyze the answers to later bracket questions, avoiding the need to model fixed-point anchoring.

Our nonparametric testing strategy is illustrated in Figure 1, which is largely derived from the numbers in Table 1. The solid line is the empirical distribution of open-ended answers. The other curves use only the answers to the first bracket question to estimate the distributions for the samples of immediate bracket respondents and follow-up bracket respondents. For example, for the immediate bracket respondents, the fraction of households with consumption less than $2000 is based upon the group of immediate respondents who got $2000 as entry point – the point (20,0.688) on the short-dashed curve in Figure 1.

The three distributions in Figure 1 differ if non-response to the open-ended question is non-random (i.e., missing values are not completely missing at random, see Little and Rubin, 2002). In the sub-sample of sample I that gave an open-ended answer, 82.6% report consumption less than $2000, compared to 77.3% in the sub-sample who gave a follow-up bracket answer. An estimate for the fraction of people with household consumption less than $2000 in the total population of interest on the basis of sample I is 0.683*0.826+0.317*0.773=0.809. Under the assumption that all answers are based upon the same initial estimates of consumption $\hat{Y}$, this estimate is consistent for the population fraction with $\hat{Y}$ less than \$2000 under any form of selection into open-ended answers or follow-up brackets. The open-ended answers (solid curve) might be negatively selected, so that follow-up bracket answers (long-dashed curve) are positively selected.

The immediate bracket responses are, because of the random assignment, drawn from the same population as the combined sample of open-ended and follow-up bracket respondents (ignoring full non-respondents – see Section 6). If all answers are based on the same initial estimates of the respondents, this implies that the distribution of immediate bracket respondents should be a weighted average of the other two distributions in the figure. Figure 1 suggests this is not the case; the point estimate for sample II is only 0.688. In the next section, we develop a formal test based upon this intuition, looking at these differences at several entry points simultaneously.

4 Testing the Fixed-Point Model of Anchoring

A major distinction between the various anchoring models discussed in section 2 is their implication for the first bracket question. The fixed-point model assumes that the entry point acts as an anchor for later bracket questions but does not affect whether the amount is larger or smaller than the entry point itself, and thus does not affect the answer to the first bracket question. The same applies to any Bayesian model for anchoring: the first bracket point leads to an update of prior information, but will not move the initial estimate – a mean or other location measure of the respondent's prior distribution – past the entry point. As explained in section 2, the gating model, on the other hand, assumes that new errors can be made at every bracket question, including the first one. This applies to the simple fixed point gating model (see section 2) as well as the more advanced model of Hurd et al. (1998). In the latter, explaining why the distribution shifts in the direction of the first entry point even requires that larger errors are made in the first question than in later unfolding bracket questions.

As shown in section 2 for the fixed probability gating model, gating reduces some probabilities (those larger than 0.5) and increases other probabilities (those smaller than 0.5). The same holds in the Hurd et al. (1998) model. On the other hand, since all the bracket questions in the experiment are of the form "Is the amount E or more?" acquiescence al-

Table 1. Distribution of Reported Consumption

Sample I: open-ended question first				
	E=500	E=1000	E=2000	All E
observations	674	698	725	2097
obs. open answer	435	485	496	1416
(percentage)	(65.8)	(71.2)	(69.2)	(68.3)
% of open answers				
< 500	17.9	16.8	17.8	17.5
<1000	48.2	49.4	51.1	49.6
<2000	80.7	79.7	82.6	81.0
obs. bracket answer	239	213	229	681
(percentage)	(34.2)	(29.8)	(30.8)	(31.7)
% of bracket answers				
<500	18.5	16.2	11.4	
<1000	63.0	52.6	42.6	
<2000	90.5	87.1	77.3	
Sample II: forced brackets				
	E=500	E=1000	E=2000	E=5000
observations	690	612	582	778
% of bracket answers				
< 500	15.9	11.6	7.0	5.4
<1000	55.7	40.9	29.9	31.0
<2000	84.1	83.7	68.8	64.2
<5000	98.2	98.9	97.5	94.2

Notes: respondents who answer no open question or bracket question are not included (cf. Appendix A). All percentages weighted with HRS provided sample weights.

ways goes in one direction: it leads to underestimation of $P[\hat{Y} < E]$ and to overestimation of $P[\hat{Y} \geq E]$, where $\hat{Y}$ is the initial estimate not affected by the format of the bracket question.

If samples of open-ended answers and bracket answers drawn randomly from the same population were available, the joint null hypothesis of no acquiescence bias and no anchoring in the first bracket question could be tested by comparing fractions with $P[\hat{Y} \geq E]$ in the open-ended answers with fractions answering "yes" to the bracket question whether the amount is E or more in the sample of bracket answers. Moreover, the nature of the deviations would indicate whether acquiescence or anchoring is the main problem. For example, at a low entry point (with $P[\hat{Y} \geq E] > 0.5$) anchoring according to the gating model would predict that the fraction "yes" in the bracket answers is lower than $P[\hat{Y} \geq E]$ in the open-ended answers, whereas yea-saying would predict the reverse.

The reality of the experiment is more complicated. First, some respondents do not reveal any information. In the sample of immediate bracket respondents (sample II), these are the respondents that do not answer the first bracket question. In the sample of those who start with an open-ended question (sample I), some people answer neither the open-ended nor the first bracket question in the follow-up unfolding brackets design. In this section we will consider the "best-case scenario" that the complete non-respondents are similar in the two sub-samples: we assume that people who do not answer an initial bracket question would also not have answered an open-ended question, and people who do not answer a follow-up bracket question would not have answered the same bracket question either if it had been preceded by an open-ended question. The population actually analyzed excludes the people who do not reveal any information at all, and we assume that both samples (I and II) are random samples from this population. Appendix A presents some information on complete non-response for each group. It suggests there are no systematic differences between samples I and II, backing up the assumption that is made. Moreover, in section 6 we show how the assumption can be relaxed and look at "worst case scenarios."

A second complication is the substantial number of respondents in sample I who do not answer the open-ended question but do answer one or more follow-up bracket questions (cf. Table 1). If giving an open-ended answer is non-random, ignoring the bracket respondents in sample I leads to selection bias. The bracket responses in sample I may suffer from similar anchoring or acquiescence biases as those in sample II. We will combine estimates of $P[\hat{Y} < E]$ for various entry point values in the sub-samples of open-ended and bracket respondents in sample I and compare them with estimates on sample II. As explained in section 3, we will not use the information in bracket questions other than the first one, which would suffer from anchoring in any anchoring model. Under our null hypothesis, answers to the first bracket question will not suffer from anchoring or yea-saying in either group, and the difference between the sample I and II based estimates should be insignificant. Under the alternative, both the sample I and the sample II based estimates will be affected by anchoring or yea-saying. Since the majority of the sample I answers are open answers, however, the anchoring or yea-saying effect is expected to be larger in sample II than in sample I. This will drive the power of the test and helps to interpret deviations from the null.

To make this more precise, first note that, as explained in section 1, we cannot say much about the distribution of true consumption amounts, since we have only one (self-reported) amount per respondent. We focus on the respondent's initial estimate $\hat{Y}_i$, which is reported if the respondent gives an open-ended answer. Let O, B and N denote the events that an open-ended answer, a bracket answer, or no answer at all is given, respectively. Let $P_{I,E}[.]$ and $P_{II,E}[.]$ denote probabilities under the sample design of sample I or sample II with entry point E. The assumption maintained in this section that full non-response is the same in the two samples, is given by:

$$(A1)\quad P_{I,E_1}[N|\hat{Y}] = P_{II,E_2}[N|\hat{Y}],\ \ E_1, E_2 = 500, 1000, 2000$$

Note that $(A1)$ also implies that $P_{I,E}[N|\hat{Y}]$ and $P_{II,E}[N|\hat{Y}]$ do not depend on E. This will be relaxed in section 6.

In the remainder of this section we consider the populations of those who would give a bracket or an open-ended answer under the questionnaire design of samples I and II. Under $(A1)$, the distribution of $\hat{Y}$ in these two populations is the same. What we want to test is whether, in these populations, answers to bracket questions are exactly based on comparing $\hat{Y}$ to the entry point E. We cannot test this type of consistency at the individual level (since there is only one observation per respondent), but we can test the implication for the probabilities:

$$P_{J,E}[\hat{Y} < E|B] = P_{J,E}[\text{“}no\text{”}|B],\ \ E = 500, 1000, 2000; J = I, II \tag{1}$$

where $P_{J,E}[\text{“}no\text{”}|B]$ is the probability that the first bracket question (“is the amount E or more?”) is answered with “no”.

If equation (1) is satisfied, then we have, for samples I and II, respectively:

$$P_{I,E}[\hat{Y} < E|O\ or\ B] = P_{I,E}[O|O\ or\ B]P_{I,E}[\hat{Y} < E|O] + P_{I,E}[B|O\ or\ B]P_{I,E}[\text{“}no\text{”}|B].$$

and

$$P_{II,E}[\hat{Y} < E|B] = P_{II,E}[\text{“}no\text{”}|B]$$

Under assumption $(A1)$, the probabilities $P_{I,E}[\hat{Y} < E|O\ or\ B]$ and $P_{II,E}[\hat{Y} < E|B]$ are equal. This leads to the following null hypothesis, implied by (1):

$$H_0:\ P_{I,E}[O|O\ or\ B]P_{I,E}[\hat{Y} < E|O] + P_{I,E}[B|O\ or\ B]P_{I,E}[\text{“}no\text{”}|B] = P_{II,E}[\text{“}no\text{”}|B] \tag{2}$$

Under the alternative (and maintaining $(A1)$), $P_{J,E}[\hat{Y} < E|B]$ will not be equal to $P_{J,E}[\text{“}no\text{”}|B]$, $J = I, II$. Since this affects only the minority of bracket respondents in sample I but everyone in sample II, the left hand side and right hand side of (2) will in general not be equal under the alternative. The attenuation by open-ended responses makes it likely that the left hand side of (2) is then closer to $P_I[\hat{Y} < E|O\,or\,B] = P_{II}[\hat{Y} < E|B]$ than the right hand side. This will help to determine the sign of the anchoring and/or acquiescence bias.

The easiest way to implement the joint test for the three entry points E_1=500, E_2=1000 and E_3=2000 is to formulate a model which has the magnitudes in (2) for the three entry points as parameters and can be estimated by pseudo maximum likelihood (not using the observations on full non-respondents and conditioning on this, and maintaining ($A1$), and accounting for the sample design with clustering and sample weights, see, e.g., Skinner, 1989). A Wald test can then be performed on the three restrictions in (2), one restriction for each entry point.

The parameters of the unrestricted model are given by:
$\rho_{I,O} = P_I[O|O\,or\,B]$, $\pi_{I,O,1} = P_I[\hat{Y} < 500|O]$, $\pi_{I,O,2} = P_I[\hat{Y} = 500|O]$, $\pi_{I,O,3} = P_I[500 < \hat{Y} < 1000|O]$, ...,$\pi_{I,O,6} = P_I[\hat{Y} = 2000|O]$;
$\pi_{I,B,j} = P_I[$"no" to "Is it E_j or more?"$|B]$,$j = 1, 2, 3$, $E_1 = 500, E_2 = 1000, E_3 = 2000$;
$\pi_{II,j} = P_{II}[$"no" to "Is it E_j or more?"$|B]$,$j = 1, 2, 3$.

$P_I[\hat{Y} > 2000|O]$ is determined by the other parameters, since probabilities have to add up to one. The three restrictions to be tested are

$$\pi_{II,j} = \rho_{I,O}\pi_{I,O,j} + (1 - \rho_{I,O})\pi_{I,B,j}, \;\; j = 1, 2, 3.$$

Here we have imposed that $P_I[O|O\,or\,B]$ is the same across the different sub-groups of sample I. This seems plausible a priori, for the same reason why assumption $A1$ is plausible: group assignment is completely random, and respondents choose whether to give an open-ended answer or not before knowing their group assignment. Moreover, while Table 1 indicates some differences in the point estimates of this response rate across groups, these differences are neither jointly nor pairwise significant at the 5% level. Still, we have also performed the test with three different parameters for the open-ended response rates by sub-group, with the same results.

The pseudo likelihood is presented in the appendix. The unrestricted ML estimates are given by the corresponding fractions in the relevant sub-samples (weighted with sample weights).

5 Results

Table 2 presents the unrestricted and restricted estimates using all continuous and bracket respondents. The log likelihood difference implies that the null hypothesis is clearly rejected at any common significance level. The bottom panel shows the implied estimates of $P_I[\hat{Y} < E|O\,or\,B]$ under the null. Comparing these in the unrestricted model with the estimates of $\pi_{II,j}$, $j = 1, \ldots, 3$, which are equal to $P_{II}[\hat{Y} < E|B]$ under the null, indicates why the null is rejected. For $E = 500$, the estimates are not significantly different (t-test statistic: 0.8). For E=1000, the sample I based estimate is significantly larger than the sample II estimate (t-value: 3.2). Since \$1000 is close to the median, it seems implausible that this is due to gating (cf. section 2), but it could be due to acquiescence bias, with too many affirmative answers to the question whether consumption is \$1000 or more. The significant difference at E=2000 (t-value: 3.0) with the same sign could be due to acquiescence bias as well as gating, since gating typically increases the probability of an "extreme" outcome (2000 or more, in this case) for bracket answers. Qualitatively, one interpretation of the results would be a combination of acquiescence and gating. At E=500, the two effects are opposite and almost cancel, apparently. For E=1000 there is not much gating, and acquiescence raises the sample II estimate more than the sample I estimate, since sample II has bracket answers only. For E=2000, gating and acquiescence reinforce each other, explaining the particularly large difference at this entry point. On the other hand, the differences between the three entry points may also be due to the fact that acquiescence is more important at higher entry points, where more respondents should say "no" and are thus vulnerable to yea-saying mistakes (cf. section 7).

According to Hurd (1999), acquiescence and anchoring become more important if respondents are more uncertain about the actual amounts. Hurd (1999) analyzes experimental data on asset holdings, and finds that the answers of the financial respondent - the person in the household who is more knowledgeable in financial matters and answers most of the financial questions - are hardly subject to acquiescence bias, while for others, acquiescence bias is quite prominent. We checked whether this also applies to consumption by considering financial respondents (almost 75% of the sample) and other respondents separately. Separate estimates for both groups are not presented to save space. Our results differ from those of Hurd (1999): The null hypothesis of no gating or acquiescence is rejected for both financial respondents and others. For both groups, the pattern is similar as for the complete sample, with sample II estimates giving smaller reported probabilities of amounts lower than the entry points. At the intermediate entry point E=\$1000, the yea-saying effect seems smaller for financial respondents than for others, with t-statistic 1.3 and 2.8, respectively. This is in line with Hurd's finding. But the t-statistics on the test at E=500 are 0.1 for financial respondents and 0.8 for others, and at E=2000, they are 3.3 and 2.0, for financial respondents and others, respectively. This suggests less clear differences between financial respondents and others, possibly because of the other age group (Hurd looks at the younger HRS cohort) or because financial respondents are better informed about asset values than others, but are not necessarily better at estimating the value of total consumption.

Table 2: Pseudo Maximum Likelihood Estimates

		Unrestricted model		Restricted model	
		Estimate	St. error	Estimate	St. error
$\rho_{I,O}$	$P_I[O\|O\,or\,B]$	0.6846	0.0168	0.6854	0.0166
$\pi_{I,O,1}$	$P_I[\hat{Y} < 500\|O]$	0.1736	0.0161	0.1646	0.0148
$\pi_{I,O,2}$	$P_I[\hat{Y} = 500\|O]$	0.0934	0.0081	0.0888	0.0078
$\pi_{I,O,3}$	$P_I[500 < \hat{Y} < 1000\|O]$	0.2300	0.0150	0.2168	0.0143
$\pi_{I,O,4}$	$P_I[\hat{Y} = 1000\|O]$	0.1433	0.0102	0.1450	0.0102
$\pi_{I,O,5}$	$P_I[1000 < \hat{Y} < 2000\|O]$	0.1708	0.0111	0.1720	0.0115
$\pi_{I,O,6}$	$P_I[\hat{Y} = 2000\|O]$	0.0733	0.0076	0.0832	0.0082
$\pi_{I,B,1}$	P_I["no" to "Is it 500 or more?"$\|B$]	0.1866	0.0291	0.1814	0.0282
$\pi_{I,B,2}$	P_I["no" to "Is it 1000 or more?"$\|B$]	0.5229	0.0416	0.4731	0.0388
$\pi_{I,B,3}$	P_I["no" to "Is it 2000 or more?"$\|B$]	0.7743	0.0290	0.7116	0.0308
$\pi_{II,1}$	P_{II}["no" to "Is it 500 or more?"$\|B$]	0.1596	0.0212	0.1699	0.0206
$\pi_{II,2}$	P_{II}["no" to "Is it 1000 or more?"$\|B$]	0.4135	0.0252	0.4711	0.0291
$\pi_{II,j}$	P_{II}["no" to "Is it 2000 or more?"$\|B$]	0.6919	0.0293	0.7634	0.0200
Implied estimates:					
$P_I[\hat{Y} < 500\|O\,or\,B]$ under $H0$		0.1777	0.0213	0.1699	0.0206
$P_I[\hat{Y} < 1000\|O\,or\,B]$ under $H0$		0.5052	0.0308	0.4711	0.0291
$P_I[\hat{Y} < 2000\|O\,or\,B]$ under $H0$		0.7996	0.0209	0.7634	0.0200
Pseudo log likelihood:			-5681.89		-5702.23
Wald test restrictions:			28.49	(p-value: 0.000)	

6 Relaxing the Assumptions about Full Non-response

Until now we have assumed that complete non-response is not selective, i.e., the respondents in sample I who neither give an open-ended nor a bracket answer have the same distribution of $\hat{Y}$ as those in sample II who do not give any bracket answer. As explained in section 2, we condition on participating in the experimental module and on not opting out before the start of the consumption questions. This gives what we called samples I and II. In sample I, 2.24% of all respondents answer "don't know" or "refuse" to both the open-ended question and the first follow-up bracket question. In sample II, 2.74% answer "don't know" or "refuse" to the first bracket question. Since the survey designs are systematically different here - an open-ended question first versus forced brackets - selection into full non-response might be different for the two samples. In this section we will investigate whether this could explain the test results in the previous section. Thus we relax assumption $(A1)$, replacing it by an "assumption" implied by the random assignment into an experimental module:

$$(A2) \quad P_{I,E_1}[\hat{Y} < E] = P_{II,E_2}[\hat{Y} < E]; \quad E_1, E_2, E = 500, 1000, 2000$$

This assumption implies that we can omit the design subscripts in the probability distribution of $\hat{Y}$. If there is no anchoring or acquiescence and (1) holds, we have, for sample design I:

$$P[\hat{Y} < E] = P_{I,E}[O]P_{I,E}[\hat{Y} < E|O] + P_{I,E}[B]P_{I,E}[\text{“}no\text{”}|B] + P_{I,E}[N]P_{I,E}[\hat{Y} < E|N].$$

Since no information is available to estimate $P_{I,E}[\hat{Y} < E|N]$, this does not lead to a consistent estimator of $P[\hat{Y} < E]$ without further assumptions. Using that $0 \leq P_{I,E}[\hat{Y} < E|N] \leq 1$, however, gives lower and upper bounds on $P[\hat{Y} < E]$ that can be consistently estimated using sample I, similar to, for example, Manski (1989, 1995). The width between these bounds is $P_{I,E}[N]$.

Similarly, for sample design II we get, if (1) is valid:

$$P[\hat{Y} < E] = P_{II,E}[B]P_{II,E}[\text{“}no\text{”}|B] + P_{II,E}[N]P_{II,E}[\hat{Y} < E|N].$$

Using that $0 \leq P_{II,E}[\hat{Y} < E|N] \leq 1$, this gives a lower bound and an upper bound on $P[\hat{Y} < E]$ that can be consistently estimated using sample II. A test of the null of no anchoring or acquiescence can be based upon comparing the two estimated bounding intervals for the parameter $P[\hat{Y} < E]$.

To avoid the complications of testing for several inequality restrictions simultaneously, we consider only one entry point at the time. Moreover, given the result of the previous section that sample I estimates of $P_{I,E}[\hat{Y} < E|O\; or\; B]$ are always larger than corresponding sample II estimates $P_{II,E}[\hat{Y} < E|B]$ (suggesting that the alternative hypothesis should be acquiescence rather than anchoring), we consider the estimate of the upper bound for sample II, $P^U_{II}[\hat{Y} < E]$ and the estimate for the lower bound for sample I, $P^L_I[\hat{Y} < E]$ and perform a one-sided test for the null hypothesis that the former is at least as large as the latter:

$$P^L_I[\hat{Y} < E] = P_{I,E}[O]P_{I,E}[\hat{Y} < E|O] + P_{I,E}[B]P_{I,E}[\text{“}no\text{”}|B].$$

$$P^U_{II}[\hat{Y} < E] = P_{II,E}[B]P_{II,E}[\text{“}no\text{”}|B] + P_{II,E}[N]$$

$$H_0: \quad P^L_I[\hat{Y} < E] \leq P^U_{II}[\hat{Y} < E]$$

This is in line with the intuition of Imbens and Manski (2004) who show that for a two-sided confidence interval of a parameter identified up to a bounding interval, one-sided significance levels can be used. To perform a t-test on the difference of the two probabilities, note that both probabilities are expectations of dummy variables observed in sub-samples (the first is a dummy with value 1 if there is an open-ended answer less than E or a bracket answer “*no*” to the question “is the amount E or more?”, observed in the sub-sample of sample I who got entry point E; the second is a dummy with value 1 if there is a bracket answer “no” to the bracket question “Is the amount E or more?” or if the first bracket question is not answered; this dummy is observed for everyone in sample II with entry point E). The estimates of these expectations are the means of these dummies over the relevant sub-samples (weighted with sample weights).

The results are summarized in Table 3. At E=500, the point estimates of the two bounding intervals overlap, and the one-sided test cannot reject the null hypothesis of no anchoring or acquiescence at E=500, a result we also obtained under assumption ($A1$). At

Table 3: Test Results Allowing for Selective Non-response

E	$P_I^L[\hat{Y} < E]$	$P_{II}^U[\hat{Y} < E]$	Difference	St. error on diff.
500	0.1774	0.1880	-0.0106	0.0287
1000	0.4915	0.4253	0.0662	0.0361
2000	0.7943	0.6973	0.0970	0.0339

Notes: Weighted subsample fractions;
Standard errors corrected for clustering

E=\$1000, the point estimates of the bounding intervals do not overlap. Even if all non-respondents in sample II have low consumption and all full non-respondents in sample I have high consumption, the estimate using sample II (0.425) remains below that for sample I (0.491). The difference is significant at the one sided 5% level (t-value 1.83). In other words, even if full non-respondents in the two samples are completely opposite groups (low versus high consumption), we can reject the null hypothesis of no anchoring or acquiescence at E=1000. Similarly, at E=\$2000, the t-value on the difference is 2.86 and the null of no anchoring or acquiescence is clearly rejected. We conclude that selective full non-response cannot explain the finding that the no anchoring or acquiescence hypothesis is rejected.

7 Fixed Probability Gating and Acquiescence

To investigate whether gating or acquiescence bias (or both) can indeed explain why the null hypothesis is rejected, we introduce a simple form of "fixed probability" gating and acquiescence bias. We maintain assumption ($A1$) in sections 4 and 5 and consider the subpopulation of respondents who provide open-ended or bracket information. First, we assume that there is a fixed fraction $P[Acq]$ of the population that will automatically answer any bracket question with "yes," irrespective of $\hat{Y}$ and E. Second, we assume that in the remainder of the population, the bracket question triggers some people to deviate from their initial estimate, and there is a fixed probability $P[Gat]$ that people give the answer which is opposite to their initial estimate ("yes" if $\hat{Y} < E$ or "no" if $\hat{Y} \geq E$). All this implies:

$$P_J[\text{“}yes\text{” to “}Is\,it\,E\,or\,more?\text{”}|B, \hat{Y} < E] = P[Acq] + (1 - P[Acq])P[Gat], \quad J = I, II$$

and

$$P_J[\text{“}no\text{” } to \text{ “}Is\, it\, E\, or\, more?\text{”}|B, \hat{Y} \geq E] = (1 - P[Acq])P[Gat]), \quad J = I, II$$

For the symmetric case with $P[Acq] = 0$, this model is essentially the same as what Copas (1988, p.234) calls “a simple model for resistant fitting.” The general case is similar to the misclassification models used by, for example, Ekholm and Palmgren (1982), Lee and Porter (1984), and Hausman et al. (1998).

The restricted model in Table 2 can be relaxed with fixed probability acquiescence and anchoring, extending it with the probabilities $P[Acq]$ and $P[Gat]$. This gives a model which is nested in the unrestricted model, which has three parameters more than the restricted model in Table 2, and thus one parameter more than the model with fixed probability acquiescence and anchoring. Table 4 presents the estimates for the latter model, as well as for the models that set either $P[Acq]$ or $P[Gat]$ to zero.

Pseudo log likelihoods are very similar for Models 1 and 3 in Table 4 and the unrestricted model in Table 2. On the other hand, all three models in Table 4 are significantly better than the restricted model in Table 2 allowing for neither acquiescence, nor gating. The probability of gating is significant at the 10% level in model 2, but as soon as acquiescence is allowed for also, gating no longer plays any role. Model 3 with a fixed probability of yea-saying and without gating is the preferred model.

This corresponds to the estimates of $P[Gat]$ and $P[Acq]$. In models 1 and 3, the estimate of $P[Acq]$ is substantial and significant. As soon as $P[Acq]$ is allowed to be non-zero, the estimate of $P[Gat]$ is very small insignificant. The results thus suggest that yea-saying plays a dominant role and explains why the null of no anchoring and no acquiescence is rejected in sections 5 and 6. Once yea-saying is accounted for in a very simple one-parameter way, there is no evidence of anchoring (or ”gating”) at the entry point.

Figure 2 compares the implications for the distribution of $\hat{Y}$ of the three models in Table 4 and the restricted “benchmark” model without gating or acquiescence in Table 2. Compared to the benchmark model, the model with gating implies a smaller dispersion in $\hat{Y}$, since some "extreme" bracket answers – "no" at E=\$500; "yes" at E=\$2000 – are explained by gating instead of low or high values of $\hat{Y}$. While the benchmark model estimates directly reflect the data, the gating model recognizes that part of the dispersion in the data is due to gating, and corrects for that. Larger differences arise when acquiescence is allowed for. Once this is done, also allowing for gating makes no difference: the two curves allowing for acquiescence are virtually the same. Both imply lower estimates of consumption than the models not allowing for acquiescence. The reason is that in the models allowing for acquiescence, some of the "yes" answers to "Is it E or more?" are explained by yea-saying rather than high values of consumption.

Table 4. Models with Fixed Acquiescence and/or Gating Probabilities

	Model 1 Gating and Acquiescence		Model 2 Gating only		Model 3 Acquiescence only	
	Est.	S.e.	Est.	S.e.	Est.	S.e.
$\rho_{I,O}$	0.6885	0.0167	0.6860	0.0168	0.6865	0.0167
$\pi_{I,O,1}$	0.1746	0.0162	0.1618	0.0148	0.1756	0.0164
$\pi_{I,O,2}$	0.0930	0.0080	0.0904	0.0079	0.0938	0.0081
$\pi_{I,O,3}$	0.2264	0.0152	0.2204	0.0151	0.2269	0.0152
$\pi_{I,O,4}$	0.1448	0.0104	0.1469	0.0105	0.1449	0.0102
$\pi_{I,O,5}$	0.1716	0.0114	0.1739	0.0114	0.1711	0.0115
$\pi_{I,O,6}$	0.0737	0.0077	0.0807	0.0087	0.0732	0.0077
$\pi_{I,B,1}$	0.2434	0.0488	0.1138	0.0552	0.2465	0.0431
$\pi_{I,B,2}$	0.6473	0.0578	0.4660	0.0439	0.6497	0.0800
$\pi_{I,B,3}$	0.9928	0.0723	0.7636	0.0538	0.9912	0.0859
$P[Gat]$	0.0023	0.0430	0.0650	0.0448	0	
$P[Acq]$	0.2144	0.0396	0		0.2144	0.0510
Implied estimates:						
$P[\hat{Y} < 500]$	0.1960	0.0335	0.1467	0.0396	0.1978	0.0295
$P[\hat{Y} < 1000]$	0.5417	0.0406	0.4705	0.0323	0.5444	0.0541
$P[\hat{Y} < 2000]$	0.8670	0.0210	0.7840	0.0207	0.8683	0.0211
Pseudo Log likelihood	-5682.37		-5700.22		-5682.44	

Some theoretical arguments (Schuman and Presser, 1996, Chapter 8) and empirical findings (e.g. Hurd, 1999) suggest that acquiescence bias should vary with the uncertainty about the quantities being queried. Table 5 shows the results of estimating model 3 in Table 4 (with acquiescence only) over a number of sub-samples. In almost all cases the estimated acquiescence probability is substantial and its 95% confidence interval does not contain the value zero. There is some variation across subgroups, and most results are in line with the notion that people who are more uncertain have a larger probability of acquiescence. For example, financial respondents and respondents with high education level are less subject to yea-saying than others. Respondents who are not the financial respondent in their household have a particularly large acquiescence probability. Using a self-reported indicator or a test-based indicator of memory quality, we find that people with better memory are less subject to yea-saying. There is only a weak relation between age or gender and the tendency of yea-saying. The income and wealth patterns seem somewhat surprising, with the higher income and higher wealth respondents more subject to yea-saying. Perhaps this is because their consumption expenditures are more diversified and harder to estimate.

8 Conclusions

We have investigated the importance of anchoring and acquiescence bias at the entry point of an unfolding bracket design. Experimental data on consumption where respondents are randomly either given an open-ended question on household consumption or immediately directed to bracket questions were used to test for the presence of these phenomena, without making specific model assumptions. The main finding is that some bias is present even at the entry point. Further analysis making additional assumptions on the nature of acquiescence and gating suggests that acquiescence bias is the main problem, while anchoring is less important. The evidence is stronger than that of Hurd et al. (1999) in the sense that we also find acquiescence among financial respondents, the most knowledgeable household members concerning financial matters. For the consumption question at hand, accounting for selective non-response but ignoring acquiescence leads to misleading conclusions on the selective nature of item non-response. It would lead to overestimation of consumption levels and underestimation of poverty rates.

Table 5. Estimated Probabilities of Acquiescence - Model 3
(Model with fixed acquiescence probability, without gating)

sub-sample	# obs.	$P[Acq]$	
		Estimate	St. error
Financial respondent	3548	0.161	0.044
Not financial respondent	1211	0.439	0.066
Borne after 1919	2088	0.233	0.061
Borne before 1920	2671	0.208	0.045
Woman	3111	0.233	0.047
Man	1648	0.195	0.110
Years education >12	1445	0.160	0.092
Years education ≤12	3314	0.208	0.040
Above median household income	2361	0.288	0.062
Below or at median household income	2397	0.152	0.040
Above median household wealth	2354	0.272	0.057
Below or at median household wealth	2365	0.145	0.041
Self-rated memory high	1494	0.083	0.071
Self-rated memory low	3265	0.274	0.036
Memory test high	1700	0.151	0.067
Memory test low	3058	0.255	0.050
All	4759	0.214	0.051

Notes:
Missing income and wealth imputed (RAND version C);
Wealth is total wealth (including housing and IRAs).
Self-rated memory: "How would you rate your memory at the present time?" Possible answers: excellent, very good, good, fair or poor. High: excellent or very good.
Memory test: number immediately recalled from a list of 40 words read to the respondent. High: at least 6 (the maximum was 10).

Reducing acquiescence bias in data collection is easy: the bracket questions can be formulated in a neutral way, asking people, for example, to choose between "more than E,", "less than E," or "about E," rather than to answer "yes" or "no". This has already been implemented in the most recent waves of, for example, the Health and Retirement Study (HRS) and AHEAD. The results of this study suggest that, with these improved questions, it seems at least safe to use the first bracket question in an unfolding bracket design at face value. Given the evidence in the literature that anchoring is a problem also in absence of acquiescence bias, this at the same time means that anchoring at follow up bracket questions must be taken seriously. The fixed point model for anchoring or another model with Bayesian learning in which answers to the later bracket questions are affected by the entry point may then be a useful tool.

References

Boyle, K.J., H.F. MacDonald, H. Cheng and D.W. McCollum (1998), Bid design and yea saying in single-bounded, dichotomous-choice questions, Land Economics, 74(1), 49-61.

Cameron, T. and J. Quiggin (1994), Estimation using contingent valuation data from a “dichotomous choice with follow-up” questionnaire, Journal of Environmental Economics and Management, 27, 218-234.

Campbell, A., P.E. Converse, W.E. Miller and D.E. Stokes (1960), The American Voter, Wiley, New York.

Copas, J.B. (1988), Binary regression models for contaminated data, Journal of the Royal Statistical Society. Series B (Methodological), 50(2), 225-253.

Couch, A. and K. Keniston (1960), Yeasayers and naysayers: Agreeing response set as a personality variable, Journal of Abnormal and Social Psychology, 60, 151-174.

Ekholm, A. and J. Palmgren (1982), A model for a binary response with misclassification, GLIM-82: Proceedings of the International Conference on Generalized Linear Models, R. Gilchrist (ed.), Springer, Heidelberg, 128-143.

Frykblom, P. and J. Shogren (2000), An experimental testing of anchoring effects in discrete choice questions, Environmental and Resource Economics, 16, 329-341.

Green, D., K. Jacowitz, D. Kahneman and D. McFadden (1998), Referendum contingent valuation, anchoring, and willingness to pay for public goods, Resource and Energy Economics 20, 85-116.

Hausman, J.A., J. Abrevaya and F.M. Scott-Morton (1998), Misclassification of the dependent variable in a discrete response setting, Journal of Econometrics, 87(2), 239-269.

Herriges, J. and J. Shogren (1996), Starting point bias in dichotomous choice valuation with follow-up questioning, Journal of Environmental Economics and Management, 30, 112-131.

Hurd, M. (1999), Anchoring and acquiescence bias in measuring assets in household surveys, Journal of Risk and Uncertainty, 19(1-3), 111-136.

Hurd, M., D. McFadden, H. Chand, L. Gan, A. Merrill, and M. Roberts (1998), Consumption and savings balances of the elderly: Experimental evidence on survey response bias, in D. Wise (ed.), Frontiers in the Economics of Aging, University of Chicago Press, Chicago, 353-387.

Hurd, M., A. Kapteyn and J. Zissimopoulos (2001), Recovering the distribution from bracketed responses, paper presented at TMR Savings and Pensions workshop in Evian, France, August 21-23.

Imbens, G. and C. Manski (2004), Confidence intervals for partially identified parameters, Econometrica, 72(6), 1845-1857.

Jacowitz, K. and D. Kahneman (1995), Measures of anchoring in estimation tasks, Personality and Social Psychology Bulletin, 21, 1161-1166.

Jackman, M. (1973), Education and prejudice or education and response set?, American Sociological Review, 38, 327-339.

Juster, T., and J.P. Smith (1997), Improving the quality of Economic data: Lessons from the HRS and AHEAD, Journal of the American Statistical Association, 92, 1268- 1278.

Lee, L.F. and R.H. Porter (1984), Switching regression models with imperfect sample information with an application on cartel stability, Econometrica, 52, 391-418.

Lenski, G.E. and J.C. Leggett (1960), Caste, class and deference in the research interview, American Journal of Sociology, 65, 463-467.

Little, R.J.A., and D.B. Rubin (2002), Statistical analysis with missing data, Wiley, New York.

Manski, C.F., (1989), Anatomy of the selection problem, Journal of Human resources, 24, 343-360.

Manski, C.F., (1995), Identification problems in the social sciences, Harvard University Press.

Northcraft, G. and M. Neale (1987), Experts, amateurs and real estate: an anchoring and adjustment perspective on property pricing decisions, Organizational Behavior and Human Decision Processes, 39, 84-97.

O'Connor, R., M. Johansson and P.-O. Johansson (1999), Stated preferences, real behaviour and anchoring: some empirical evidence, Environmental and Resource Economics, 13, 235-248.

Rabin, M. (1998), Psychology and economics, Journal of Economic Literature, 36(1), 11-46.

Schuman, H. and S. Presser (1996), Questions and answers in attitude surveys, SAGE Publications, Thousand Oaks.

Schwarz, N., H.-J. Hippler, B. Deutsch and F. Strack (1985), Response scales: effects of category range on reported behavior and comparative judgments, The Public Opinion Quarterly, 49(3), 388-395.

Skinner, C.J. (1989), Domain means, regression and multivariate analysis, in Skinner, C.J., D. Holt and T.M.F. Smith (eds.), Analysis of complex surveys, Wiley, Chichester, 59-88.

Strack, F. and T. Mussweiler (1997), Explaining the enigmatic anchoring effect: mechanisms of selective accessibility, Journal of Personality and Social Psychology, 73(3), 437-446.

Sudman, S., N.M. Bradburn and N. Schwarz (1996), Thinking about answers - The application of cognitive processes to survey methodology, Jossey-Bass Publishers, San Francisco.

Tversky, A. and D. Kahneman (1974), Judgment under uncertainty: heuristics and biases, Science, 185(4157), 1124-1131.

Appendix A: Details on Complete Non-Response

Table A1 gives an overview of complete item non-response in the consumption question. It concerns the 4885 respondents who were willing to answer the questions in the experimental module and did not interrupt the interview before reaching the first consumption question. Those who were not willing (1173 observations) or interrupted the interview (43 observations) are excluded.

Table A1. Complete Non-response Rates By Experimental Module

Module	Number of observations	Non-response rate
Open-ended, E=500	693	2.58%
Open-ended, E=1000	714	2.23%
Open-ended, E=2000	742	1.95%
Forced bracket, E=500	709	3.45%
Forced bracket, E=1000	631	2.71%
Forced bracket, E=2000	602	3.08%
Forced bracket, E=5000	795	1.85%
Total	4885	2.52%

Notes: Percentages weighted with sample weights.
"Open-ended" respondents first get an open-ended question and go into brackets if they do not answer this; "Forced bracket" respondents only get bracket questions.
Non-response rate: those who do not answer any open-ended or bracket question.

Non-response rates vary from 1.95% to 3.45%. The differences between the seven groups are insignificant (the design-based Pearson F-test on independence has a p-value of 0.582). The difference between those who start with an open-ended question and those who start with a bracket question is insignificant also (p-value 0.290), so that there is no evidence that complete non-response is related to the question format, supporting assumption A1 in section 4.

We have also checked whether complete non-response is correlated with some household or individual characteristics. We found no relation between non-response and, for example, age, education level, gender, race, or marital status. We found that financial respondents, i.e., the household members most knowledgeable in financial matters and answering the household income and household assets questions in the core interview, had significantly lower non-response (1.97%) than other household members (4.29%).

Appendix B: Likelihood Contributions

This appendix presents the pseudo likelihood contributions for the model in section 4. They are pseudo likelihood contributions in the sense that clustering is not taken into account, the maximized log likelihood is the weighted sum of contributions of individual observations, see Skinner (1989). Full non-respondents are discarded. Likelihood contributions vary with the design (sub-sample I or sub-sample II; entry point E_1=500, E_2=1000 or E_3=2000, and the outcomes: open-ended answer or bracket answer; category containing the open-ended answer; "no" or "yes" to the first bracket question. The possible likelihood contributions are listed below.

Table A1. Likelihood Contributions Unrestricted Model Section 4

Sub-sample I, open-ended answer with amount <500:	$\rho_{I,O}\pi_{I,O,1}$
Sub-sample I, open-ended answer with amount 500	$\rho_{I,O}\pi_{I,O,2}$
Sub-sample I, open-ended answer with amount >500 and < 1000	$\rho_{I,O}\pi_{I,O,3}$
Sub-sample I, open-ended answer with amount 1000	$\rho_{I,O}\pi_{I,O,4}$
Sub-sample I, open-ended answer with amount >1000 and < 2000	$\rho_{I,O}\pi_{I,O,5}$
Sub-sample I, open-ended answer with amount 2000	$\rho_{I,O}\pi_{I,O,6}$
Sub-sample I, open-ended answer with amount >2000	$\rho_{I,O}(1-\sum_{k=1}^{6}\pi_{I,O,k})$
Sub-sample I, entry point E_j, bracket answer "no", $j=1,2,3$	$(1-\rho_{I,O})\pi_{I,B,j}$
Sub-sample I, entry point E_j, bracket answer "yes", $j=1,2,3$	$(1-\rho_{I,O})(1-\pi_{I,B,j})$
Sub-sample II, entry point E_j, bracket answer "no", $j=1,2,3$	$\pi_{II,j}$
Sub-sample II, entry point E_j, bracket answer "yes", $j=1,2,3$	$(1-\pi_{II,j})$